



# NLP Solutions as Asymptotic Values of ODE Trajectories

Mazen Alamir

## ► To cite this version:

Mazen Alamir. NLP Solutions as Asymptotic Values of ODE Trajectories. [Research Report] GIPSA-lab. 2014. hal-01104686

**HAL Id: hal-01104686**

**<https://hal.science/hal-01104686>**

Submitted on 19 Jan 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# NLP Solutions as Asymptotic Values of ODE Trajectories

Mazen Alamir

CNRS/Gipsa-lab Control systems department.

Email : mazen.alamir@grenoble-inp.fr

## Abstract

In this paper, it is shown that the solutions of general differentiable constrained optimization problems can be viewed as asymptotic solutions to sets of Ordinary Differential Equations (ODEs). The construction of the ODE associated to the optimization problem is based on an exact penalty formulation in which the weighting parameter dynamics is coordinated with that of the decision variable so that there is no need to solve a sequence of optimization problems, instead, a single ODE has to be solved using available efficient methods. Examples are given in order to illustrate the results. This includes a novel systematic approach to solve combinatoric optimization problems as well as fast computation of a class of optimization problems using analogic circuits leading to fast, parallel and highly scalable solutions.

## 1 Introduction

Consider the following optimization problem with inequality constraints:

$$\min_{x \in \mathbb{R}^n} f(x) \text{ under } c_i(x) \leq 0 \quad i \in \{1, \dots, n_c\} \quad (1)$$

where  $f$  and  $c_i$ 's are scalar functions of the decision variable  $x \in \mathbb{R}^n$ . Let  $f$  be the exact penalty induced function defined by:

$$\bar{f}(x, \rho) = f(x) + \rho \cdot \psi(x) \quad (2)$$

where  $\psi(x)$  is given by:

$$\psi(x) := \sum_{i=1}^{n_c} [\max\{0, c_i(x)\}]^m \quad ; \quad m \in \mathbb{N} \quad (3)$$

A wide class of algorithms intends to solve (1) by solving a sequence of unconstrained optimization problems of the form

$$\min_{x \in \mathbb{R}^n} \bar{f}(x, \rho_k) \quad (4)$$

for a varying (generally increasing) values of the weighting coefficient  $\rho_k$ . For each problem in the sequence, only  $x$  is searched for while  $\rho_k$  is kept constant [2, 1]. The series of unconstrained problems are sometimes replaced a series of problems with by box constraints as descent method with projection are easy to perform.

The increase of  $\rho$  is generally defined by  $\rho_{k+1} \leftarrow r \times \rho_k$  with  $r > 1$ . The sequence of precision parameter  $\omega_k$  (to which the intermediate problems have to be solved) is also made such that  $\omega_k \rightarrow 0$  in order to avoid solving with a uselessly high precision intermediate problems. It is then obvious that the efficiency of the resulting algorithms is tightly related to the choice of  $r$  and the intermediate precision sequence  $\{\omega_k\}_k$  since small values of  $r > 1$  leads to unnecessarily high number of intermediate problems while a too high values of  $r > 1$  leads to stiff problems that may lead to slow convergence (because this makes the solution of the box-constraint sub-problem harder [1]) beside the fact that it breaks the *continuation* argument that underlies the whole scheme. The choice of  $\omega_k$  corresponds to similar trade-offs that need to be carefully handled. Such issues are extensively studied in [1] leading to non necessarily monotonic behavior of  $\rho_k$  when solving the sequence of intermediate box constrained modified Lagrangian problems in order to avoid high number of iterations that result when  $\rho_k$

is unnecessarily high. This recent study [1] shows at least that monitoring the dynamic evolution of  $\rho_k$  is not a trivial issue.

In this paper, it is shown that simultaneous dynamics of  $x$  and  $\rho$  can be defined through a differential equation of the form:

$$\dot{x} = F_1(x, \rho) ; \dot{\rho} = F_2(x, \rho) \quad (5)$$

such that solving (5) gives trajectories that asymptotically converge towards the set of solutions of (1).

Note that by doing so, the present paper does not propose a specific alternative algorithm to solve the NLP problem (1). Rather, it enables all the efficient algorithms that are available through the huge literature on ODE integration to become candidate algorithms for (1). Moreover all the computational background regarding many issues (such as parametric sensitivity [7], parallel computing [13], precision monitoring to cite but few items) become available for the constrained optimization problem paradigm. As such, the result of the present paper can be viewed as a starting point for future investigation. More interestingly, it is shown briefly in this paper that expressing the fact that solving (1) can be done by integrating ODEs enables (for some specific problems) to build analogic circuits that can achieve the task in fast, parallel and massively scalable way.

This paper is organized as follows: first section 1.1 gives the definitions and notation used throughout the paper. Section 2 states the working assumptions that are needed to derive the main results of the paper. These results are stated and proved in section 3 while section 4 gives some examples of application of the paper results. Finally the paper ends with section 5 that summarizes the contribution and gives hints for further investigations.

## 1.1 Definition & Notation

Throughout the paper, the following notation is used. The  $n$ -dimensional vectors  $f_x(x)$ ,  $\psi_x(x)$  and  $\bar{f}_x(x, \rho)$  denote the gradients of  $f$ ,  $\psi$  and  $\bar{f}$  w.r.t  $x$ . The scalar function  $g(x, \rho)$  denotes the euclidian norm of  $\bar{f}_x$  at  $(x, \rho)$ , namely:

$$g(x, \rho) := \|\bar{f}_x(x, \rho)\| \quad (6)$$

For a given weighting coefficient  $\rho > 0$ , the set of stationary points of  $\bar{f}(\cdot, \rho)$  is denoted by  $\mathcal{S}_\rho$ , namely:

$$\mathcal{S}_\rho := \{x \in \mathbb{R}^n \mid g(x, \rho) = 0\} \quad (7)$$

For any  $x \in \mathbb{R}^n$ , the notation  $d(x, \rho)$  refer to the distance between  $x$  and the set  $\mathcal{S}_\rho$ , namely:

$$d(x, \rho) := \min_{z \in \mathcal{S}_\rho} \|x - z\| \quad (8)$$

The set of admissible values of  $x$  is denoted by:

$$\mathcal{A} := \{x \in \mathbb{R}^n \mid \psi(x) = 0\} \quad (9)$$

## 2 Working Assumptions

The first assumption states that the optimization problem is well posed in the sense that either the original cost  $f(x)$  is already lower bounded or the constraints are such that the weighted cost  $\bar{f}$  is lower bounded:

**Assumption 1 [Well posedness]** *For any  $\rho > 0$ , there is a lower bound  $\bar{f}_{\min}(\rho)$  such that  $\bar{f}(x, \rho) \geq \bar{f}_{\min}(\rho)$  for all  $x \in \mathbb{R}^n$ .*

Note that in the framework of the present paper, it is not assumed that the functions involved are convex. This means that the set  $\mathcal{S}_\rho$  may not be a singleton  $\{x^*(\rho)\}$ , it is assumed that the norm of the gradient of the weighted cost  $\bar{f}(\cdot, \rho)$  **away** from  $\mathcal{S}_\rho$  can be bounded below by the distance to the set  $\mathcal{S}_\rho$  through some coefficient  $k_c$ . This leads to the following generalization of the strong convexity assumption:

**Assumption 2 [( $S_\rho$ )-Strong Convexity]** *There is a constant  $k_c > 0$  such that the following inequality holds:*

$$g(x, \rho) \geq k_c \times d(x, \rho) \quad (10)$$

for all  $x \in \mathbb{R}^n$ .

Note that contrary to the classical strong convexity assumption that involves two arbitrary points  $x_1$  and  $x_2$ , the inequality (10) involves the distance from an arbitrary  $x$  to those points lying inside the set of stationary points  $\mathcal{S}_\rho$ .

The Next assumption describes a generalized Lipschitz-like assumption on the constraints and the way they are used to construct the exact penalty term  $\psi(x)$ .

**Assumption 3 [Growth rate of  $\psi$ ]** *There is a polynomial  $P$  of degree  $n_\psi \in \mathbb{N}$  with  $P(0) = 0$  that satisfies the following inequality*

$$|\psi(x_2) - \psi(x_1)| \leq P(\|x_2 - x_1\|) \quad (11)$$

for all  $x \in \mathbb{R}^n$ .

Note that the use of the polynomial  $P$  of the form:

$$P(d) := \sum_{i=1}^{n_\psi} \alpha_i d^i \quad (12)$$

accounts for the possibility to use different penalty exponents  $m$  in the definition of the constraint penalty term in (3) and the fact that the bounding function may involve lower powers for small distances  $d$  and higher powers far from the set  $\mathcal{A}$ .

The following assumption is needed to guarantee the existence of solutions to the ODE built up with the functions  $\bar{f}_x$  and  $\psi$ :

**Assumption 4 [Locally-Lipschitz maps]** *For all finite  $\rho > 0$  the maps  $f_x(\cdot, \rho)$ ,  $\psi(\cdot)$  and  $\psi_x(\cdot)$  are locally Lipschitz.*

Note that this last assumption expresses *local* requirement while (10) and (11) are required to hold for any  $x$ .

The last assumption concerns the relevance of the use of the penalty method to solve (1). It states that when the penalty coefficient  $\rho$  goes to infinity, the possible stationary points for the weighted cost converge towards the admissible set  $\mathcal{A}$ :

**Assumption 5 [Relevance of the penalty approach]**

$$\lim_{\rho \rightarrow \infty} \left[ \sup_{x \in \mathcal{S}_\rho} \psi(x) \right] = 0 \quad (13)$$

This assumption is almost implicitly required in any penalty-based approach to solve the constrained optimization problem (1). It can obviously be replaced by some more apparently trivial assumptions that can be used to prove (13). The short form is preferred here for the sake of clarity.

### 3 Main Results

The main result of the paper can be stated in the following proposition:

**Proposition 1 [Main Result]** Assume that some  $(\lambda, q) \in \mathbb{R}_+^* \times \mathbb{N}$  is chosen. Consider the following system of differential equations:

$$\dot{x} = - \left[ \sum_{i=1}^q \frac{(\lambda \cdot g(x, \rho))^{i-1}}{(i-1)!} \right] \times \bar{f}_x(x, \rho) \quad (14)$$

$$\dot{\rho} = \gamma \times \psi(x) \quad (15)$$

if the following conditions hold:

1. Assumptions 1-5 are satisfied

2.  $q \geq n_\psi$  [see (12)]

then for any  $\lambda > 0$ , there is a sufficiently small  $\gamma > 0$  such that any asymptotic solution of (14)-(15) satisfies the KKT necessary conditions of optimality for the constrained optimization problem (1).  $\heartsuit$

PROOF. Let us compute the derivative of the weighted cost  $\bar{f}(x, \rho)$ :

$$\begin{aligned} \frac{d\bar{f}}{dt} &= \gamma \psi^2(x) - \left[ \sum_{i=1}^q \frac{(\lambda \cdot g(x, \rho))^{i-1}}{(i-1)!} \right]^2 \times g^2(x, \rho) \\ &= \gamma \psi^2(x) - \left[ \sum_{i=1}^q \frac{(\lambda \cdot g(x, \rho))^i}{\lambda \cdot (i-1)!} \right]^2 \end{aligned} \quad (16)$$

Let  $x_s(\rho) \in \mathcal{S}_\rho$  be the closest point to  $x$  that lies inside the stationary set  $\mathcal{S}_\rho$ . According to (11) of Assumption 3, one can write:

$$\psi(x) \leq \psi(x_s(\rho)) + P(\|x - x_s(\rho)\|) \quad (17)$$

$$\leq \psi(x_s(\rho)) + P(d(x, \rho)) \quad (18)$$

Now by virtue of Assumption 5,  $\psi(x_s(\rho))$  satisfies the following asymptotic property:

$$\psi(x_s(\rho)) = O(1/\rho) \quad (19)$$

Therefore, (18) becomes [using (12)] :

$$\psi(x) \leq P(d(x, \rho)) + O(1/\rho) \quad (20)$$

$$\leq \left[ \sum_{i=1}^{n_\psi} \alpha_i d^i(x, \rho) \right] + O(1/\rho) \quad (21)$$

On the other hand, the sum in the r.h.s of (16) satisfies [because of (10)] the following inequality:

$$\left[ \sum_{i=1}^q \frac{(\lambda \cdot g(x, \rho))^i}{\lambda \cdot (i-1)!} \right]^2 \geq \left[ \sum_{i=1}^q \frac{(\lambda \cdot k_c \cdot d(x, \rho))^i}{\lambda \cdot (i-1)!} \right]^2 \quad (22)$$

Now using (21) and (22) in (16) enables to write [dropping all the terms with indices higher than  $n_\psi \leq q$  in the summing term of (16)] and using the identity  $y_1^2 - y_2^2 = (y_1 + y_2)(y_2 - y_1)$ :

$$\frac{d\bar{f}}{dt} \leq \left[ \sum_{i=1}^{n_\psi} \beta_i^+ d^i + O(1/\rho) \right] \cdot \left[ \sum_{i=1}^{n_\psi} \beta_i^- d^i + O(1/\rho) \right] \quad (23)$$

where  $d^i := d^i(x, \rho)$  while  $\beta_i^+$  and  $\beta_i^-$  are given by:

$$\beta_i^+ = \sqrt{\gamma} \alpha_i + \frac{(\lambda k_c)^i}{\lambda \cdot (i-1)!} \quad (24)$$

$$\beta_i^- = \sqrt{\gamma} \alpha_i - \frac{(\lambda k_c)^i}{\lambda \cdot (i-1)!} \quad (25)$$

and taking  $\gamma$  sufficiently small so as to satisfy the following inequality:

$$\sqrt{\gamma} \leq \min_{i=1}^{n_\psi} \left[ \frac{(\lambda k_c)^i}{2\alpha_i(\lambda \cdot (i-1)!)} \right] \quad (26)$$

the following inequalities hold for  $\beta_i^+$  and  $\beta_i^-$ :

$$\beta_i^+ \geq \frac{(\lambda k_c)^i}{\lambda \cdot (i-1)!} \quad ; \quad \beta_i^- \leq -\frac{(\lambda k_c)^i}{2\lambda \cdot \alpha_i((i-1)!)} \quad (27)$$

With these inequalities, inequality (23) implies:

$$\begin{aligned} \frac{d\bar{f}}{dt} \leq & - \left[ \sum_{i=1}^{n_\psi} \frac{(\lambda k_c)^i}{(\lambda \cdot (i-1)!)} d^i(x, \rho) + O(1/\rho) \right] \times \\ & \left[ \sum_{i=1}^{n_\psi} \frac{(\lambda k_c)^i}{2(\lambda \cdot (i-1)!)} d^i(x, \rho) + O(1/\rho) \right] \end{aligned} \quad (28)$$

Let us now show that inequality (28) implies that  $\lim_{t \rightarrow \infty} \psi(x(t)) = 0$ . Indeed, if this was not the case, then by the very definition of the dynamic on  $\rho$  [see (15)] it comes that  $\rho$  goes to infinity. This together with (28) and the lower boundedness of  $\bar{f}$  [Assumption 1] implies that  $\lim_{t \rightarrow \infty} d(x, \rho) = 0$  ( $x$  converges to the set  $\mathcal{S}_\rho$ ). But this implies by (13) of assumption 5 that  $\psi(x)$  converges to 0 which contradicts the assumption. Now since  $\psi(x)$  converges to 0, the inequality (16) together with the lower boundedness of  $\bar{f}$  implies also that  $g(x, \rho)$  converges to 0.

By now it has been shown that provided that  $\gamma$  is sufficiently small to satisfy (26), the trajectory of  $(x, \rho)$  converges to the following set

$$\left\{ (x, \rho) \mid g(x, \rho) = 0 \text{ and } \psi(x) = 0 \right\} \quad (29)$$

It remains to prove that if  $(x, \rho)$  belongs to the set defined by (29), then  $x$  satisfies the KKT necessary conditions of optimality. Remember that these conditions require the existence of a vector  $\mu \in \mathbb{R}^{n_c}$  such that the following conditions hold:

$$\begin{aligned} f_x(x) + \sum_{i=1}^{n_c} \mu_i \frac{\partial c_i}{\partial x}(x) &= 0 \\ c_i(x) &\leq 0 & (\forall i \in \{1, \dots, n_c\}) \\ \mu_i &\geq 0 & (\forall i \in \{1, \dots, n_c\}) \\ \mu_i \times c_i(x) &= 0 & (\forall i \in \{1, \dots, n_c\}) \end{aligned} \quad (30)$$

But  $g(x, \rho) = 0$  can be explicitly written as follows:

$$f_x(x) + \rho \times m \sum_{i=1}^{n_c} [\max\{0, c_i(x)\}]^{m-1} \times \frac{\partial c_i}{\partial x}(x) \quad (31)$$

which obviously shows that by taking  $\mu$  such that:

$$\mu_i := \begin{cases} 0 & \text{if } c_i(x) < 0 \\ \rho \times m \times [c_i(x)]^{m-1} & \text{if } c_i(x) \geq 0 \end{cases} \quad (32)$$

the first KKT condition is satisfied by construction. The second condition ( $c_i(x) \leq 0$ ) results from  $\psi(x) = 0$ . The third and the fifth conditions ( $\mu_i \geq 0$  and  $\mu_i \cdot c_i(x) = 0$ ) result from (32). This ends the proof.  $\square$

Note that if the summation in (14) is performed with an infinite number of terms, the following corollary can be obtained:

**Corollary 3.1** *Assume that some  $\lambda > 0$  is chosen. Consider the following system of differential equation:*

$$\dot{x} = -\exp[\lambda \cdot g(x, \rho)] \times \bar{f}_x(x, \rho) \quad (33)$$

$$\dot{\rho} = \gamma \times \psi(x) \quad (34)$$

**If Assumptions 1-5 hold then for sufficiently small  $\gamma > 0$ , any asymptotic solution of (33)-(34) satisfies the necessary conditions of optimality for the constrained optimization problem (1).**  $\heartsuit$

Note that the result of Proposition 1 holds for any initial condition that can be used to initialize the trajectory of (14)-(15). The price to obtain such a global result lies in the use of the  $q$ -term summation that premultiplies the gradient term  $-\bar{f}_x(x, \rho)$  in (14). The next proposition gives a weaker result that can nevertheless be preferable in some circumstances. In this weaker result, the convenient sufficiently small  $\gamma$  would depend on the initial values of  $x$  and  $\rho$ .

**Proposition 2 [A Simpler Weaker Result]** *Assume that some  $\lambda > 0$  is chosen. Consider the following system of differential equations:*

$$\dot{x} = -\bar{f}_x(x, \rho) \quad (35)$$

$$\dot{\rho} = \gamma \times \psi(x) \quad (36)$$

*If the following conditions hold:*

1. *Assumptions 1-5 are satisfied,*
2.  *$f(\cdot)$  is proper (that is  $\lim_{\|x\| \rightarrow \infty} f(x) = \infty$ )*

*then for any initialization  $(x_0, \rho_0)$ , there is sufficiently small  $\gamma > 0$  such that the resulting asymptotic solution of (35)-(36) satisfies the KKT necessary conditions of optimality for the optimization problem (1).  $\heartsuit$*

PROOF. Note that the result can be obtained if one can show that everything behaves as if  $n_\psi = 1$  holds in (11). Indeed, in this case  $q = n_\psi = 1$  can be used and (14) is equivalent to (35). This can be done using classical arguments that are typically used to derive semi-global results. More precisely, given the initial state  $(x_0, \rho_0)$ , define the following level set in  $\mathbb{R}^n$ :

$$\mathcal{V}(x_0, \rho_0) := \left\{ x \mid f(x) \leq 2\bar{f}(x_0, \rho_0) \right\} \quad (37)$$

to which the initial value  $x_0$  obviously belongs [because  $f(x_0) \leq \bar{f}(x_0, \rho)$  for all  $\rho$ ]. Note that since  $f$  is proper by assumption, the set  $\mathcal{V}(x_0, \rho_0)$  is a compact set. Consequently, there is some sufficiently high  $\bar{\alpha}_1$  such that the following inequality holds for all  $(x_1, x_2) \in \mathcal{V}(x_0, \rho_0)$ :

$$\|\psi(x_2) - \psi(x_1)\| \leq P(\|x_2 - x_1\|) \leq \bar{\alpha}_1 \|x_2 - x_1\| \quad (38)$$

this means that as far as the trajectory remains in  $\mathcal{V}(x_0, \rho_0)$ , the result of Proposition 1 can be used with  $n_\psi = 1$  therefore, there exists sufficiently small  $\gamma$  such that the dynamics defined by (14) with  $q = 1$  [which is the same as (35)] decreases the value of  $\bar{f}(x, \rho)$ . But this guarantees that the trajectory of  $x$  remains in  $\mathcal{V}(x_0, \rho_0)$ . This implies that the inequality (38) remains true and the result obviously follows.  $\square$

Note that such finite  $\bar{\alpha}_1 > 0$  exists as long as the last inequality in (38) is required only on the compact set  $\mathcal{V}(x_0, \rho_0)$ . The latter is defined in terms of the initial paire  $(x_0, \rho_0)$ . This is why the value of  $\bar{\alpha}_1$  does depend on the initialization and may not exit globally.

### 3.1 General Comments

Before getting to the examples section, it is worth mentioning that the results of the present section build a theoretical bridge between NLP and ODE algorithms in a rather systematic way and for a large class of problems. However, it must be underlined that although the following examples show rather efficient computational results, the integration of the resulting ODE may not be the more efficient way to solve the underlying optimization problems. This is because integration schemes try to reproduce high precision solution over the whole trajectories while from the NLP solution point of view, only the asymptotic trajectory matters.

To this respect, the results of the present section can be used to derive gradient-based algorithms (fast gradient for instance [9, 10]) using the r.h.s of the ODE as extended gradient in the extended space of  $(x, \rho)$  with  $\bar{f}$  as cost function. By doing so, even certification results similar to the one proposed in [12] can be extended from the case where only saturations on the control input is used to the more general case of affine constraints on the

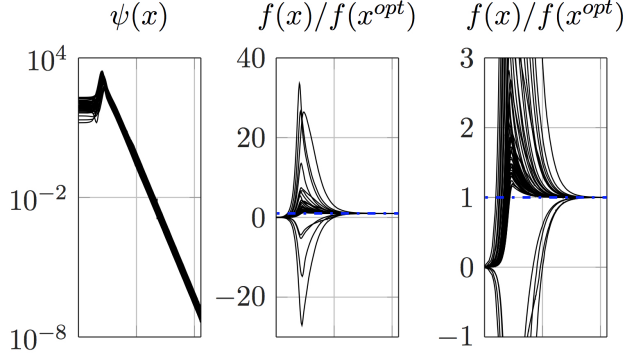


Figure 1: **Example 1.** Trajectories of  $\psi(x)$  and  $f(x)$  for 50 different randomly generated problems. Note that the constraints are asymptotically satisfied ( $\psi(x) \rightarrow 0$ ) and that the resulting costs converge towards the exact optimal value  $f(x^{opt})$  (since  $f/f^{opt} \rightarrow 1$ ). For all trials, the initial conditions  $(0,0)$  is used. Initial values of  $\psi$  shows initial strong violation of the constraints.

state. This being said, no such efficiency-oriented development is done here focusing on the main theoretical contribution of the paper.

On the other hand, another consequence of the theoretical result of the present section is the possibility to built electronic circuits that realize analogic ultra-fast integration of the ODE for a class of NLPs. This is briefly discussed in section 4.2.

## 4 Illustrative Examples

### 4.1 Example 1: QP problems

As a first examples let us consider the use of the ODE framework described in Proposition 1 to solve Quadratic Programming (QP) problem with inequality constraints. This leads to the following instantiation of the cost function  $f(x)$  and the constraints  $c_i(x)$ :

$$f(x) = \frac{1}{2}x^T Hx + F^T x \quad (39)$$

$$c_i(x) = A_i x - B_i \quad ; \quad i = 1, \dots, n_c \quad (40)$$

where  $H \in \mathbb{R}^{n \times n}$ ,  $F \in \mathbb{R}^{n \times 1}$ ,  $A_i \in \mathbb{R}^{1 \times n}$  and  $B_i \in \mathbb{R}$ . Now using  $m = 2$  to define the constraints-related weighting term:

$$\psi(x) = \sum_{i=1}^{n_c} [\max\{0, A_i x - B_i\}]^2 \quad (41)$$

gives  $n_\psi = 2$  [see (11) and (12)]. Consequently, following Proposition 1, the following ODE is defined (taking  $q = 2$ ):

$$\dot{x} = -\left[1 + \lambda \cdot \|\bar{f}_x(x, \rho)\|\right] \times \bar{f}_x(x, \rho) \quad (42)$$

$$\dot{\rho} = \gamma \times \sum_{i=1}^{n_c} [\max\{0, A_i x - B_i\}]^2 \quad (43)$$

where

$$\bar{f}_x(x, \rho) = Hx + F + \rho \sum_{i=1}^{n_c} \max\{0, A_i x - B_i\} \times A_i^T$$



Evolution of  $x_i, i = 1, \dots, 15$

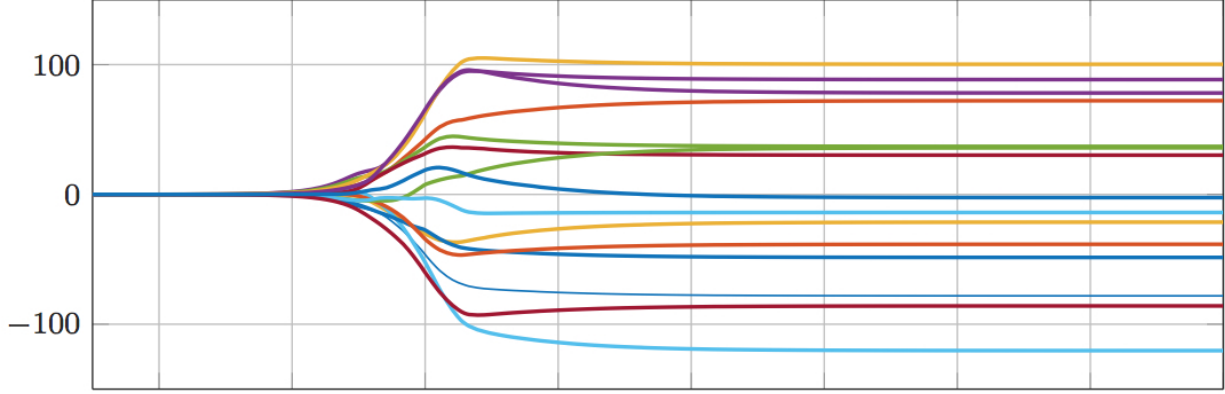


Figure 2: **Example 1.** Typical behavior of the components of  $x$  on the trajectories of the ODE. Initial value  $(0, 0) \in \mathbb{R}^n \times \mathbb{R}$  is used.

This ODE is then integrated using the Matlab ODE15s stiff solver to get the solution of the original QP problem defined by (39)-(40). Fifty randomly generated sets of matrices  $\{H, F, A, B\}$  are generated leading to 50 feasible QPs with  $n = 15$  unknown and  $n_c = 20$  constraints. The resulting ODEs (42)-(43) are defined with the parameters  $\lambda = 10^{-4}$  and  $\gamma = 10^{-6}$ . Figure 1 shows the resulting trajectories of  $\psi$  and the cost function normalized by the optimal cost value (computed using the standard Matlab QuadProg solver). All the trajectories are started from  $x_0 = 0$  and  $\rho = 0$ . The Figure clearly shows that the trajectories converge to the solutions of the problems as the constraints are satisfied and the cost function values converge toward the optimal values for all the generated problems. Figure 2 shows a typical behavior of the system's trajectory starting from  $(0, 0)$  and converging towards the optimal values. The computation times shows a mean of 49 ms with a variance of 3 ms (Using Matlab on Mac PowerBook OSX, 2.8 GHz Intel Core i7 processor).

## 4.2 Example 2: Analogic MPC Solvers

Recall that Model Predictive Control (MPC) for linear time invariant systems of the form:

$$\dot{\xi} = A\xi + Bu \quad (44)$$

is based on the repetitive solution of a quadratic programming problem of the form:

$$\min_{x \in \mathbb{R}^n} \left[ \frac{1}{2} x^T H x + [f_0 + F_1 \xi]^T x \right] \quad (45)$$

under the constraint:

$$A_i x - [b_i^0 + B_i \xi] \leq 0 \quad i \in \{1, \dots, n_c\} \quad (46)$$

where  $x \in \mathbb{R}^n$  is the parameter vector that defines the control trajectory over the prediction horizon, namely:

$$\begin{pmatrix} u(k) \\ \vdots \\ u(k+N-1) \end{pmatrix} = \Pi \cdot x \quad (47)$$

for some appropriately chosen parametrization matrix  $\Pi \in \mathbb{R}^{(Nn_u) \times n}$  where  $n_u$  is the dimension of the control input  $u$ . Note that the only difference between (39)-(40) and (45)-(46) is that the affine term in (45) and the r.h.s of the inequalities (46) depends on the state of the controlled system  $\xi$ . For more details on MPC design, the reader can refer to [8]

Now applying Corollary 3.1 to the QP defined by (45)-(46) with a sufficiently small  $\gamma$  for all initial conditions of interest, it comes that the QP solution (for a given  $\xi$ ) can be obtained by integrating the following set of ODEs:

$$\begin{aligned}\dot{x} &= -[Hx + f_0 + F\xi] + \\ &+ 2\rho \sum_{i=1}^{n_c} [\max\{0, A_i x - b_i^0 - B_i \xi\}] \cdot A_i^T\end{aligned}\tag{48}$$

$$\dot{\rho} = \gamma \sum_{i=1}^{n_c} [\max\{0, A_i x - b_i^0 - B_i \xi\}]^2\tag{49}$$

The idea is then to perform the integration through analogic circuits. The literature is very rich regarding the way transfer functions and more generally nonlinear differential relationships can be realized by analogic circuits (see [11, 5] and the references therein). Let us concentrate on the operations involved in (48)-(49) to check that analogic realizations can be derived considering that  $x$  is represented by a vector of currents (voltage options is also possible [3] although it is not discussed here):

- *Constant current sources* Note first of all that the constant terms  $f_0$  and  $b_i^0$  corresponds to tunable source of currents.
- *State dependent current sources* The state dependent terms  $F\xi$  and  $B_i\xi$  are computed numerically and the results is assigned to another vector of current sources that remain constant during the integration step. This is the only numerical operation which determines the sampling rate of the resulting MPC controller.
- *Linear combination of currents.* This concerns the terms  $Hx$  and  $A_i x$  and can be realized for instance using unity gain cells as shown for instance in [5].
- *Current summation and subtraction.* This concerns the realization of the sums  $Hx + f_0 + F\xi$  and  $A_i x - b_i^0 - B_i \xi$  and can be viewed as a particular instantiation of the previous item and can therefore be realized using unity gain cells.
- *Squaring signals.* This is necessary to compute the summated terms in (49) and can be achieved for instance using the circuits proposed in [4, 6] or any later work containing more recent devices and architectures.
- *Multiplication by  $\rho$ .* This operation can be realized through tunable gain or by using standard signal multipliers as the on proposed in [6].

Note that the time needed to analogically integrate (48)-(49) is the time necessary to fill the corresponding circuit's capacitors. This time can be made extremely short (nano or even pico-seconds) if the problem is appropriately normalized so as to have its normalized solution components  $\bar{x}_i$  scaled down so that they correspond to tiny capacitor voltages.

Note that in the above presentation, the linear character of the controlled system plays no determinant role. Indeed, thanks to the possibility of signal multiplication, squaring and even the possibility to implement the square rooting of signals [4], a wide class of ODE's that would be associated to the solution of a wide class of non quadratic constrained NLP can be analogically solved in extremely fast way. Moreover, the potential use of massively integrated circuit makes it possible to solve large scale problems in this way.

Note finally that many of the above mentioned circuits can be realized using on-line assignable gains which makes it potentially possible to use the same circuits for many different problems. It remains however necessary to analyse the cost of such circuit design and realization which is beyond the scope of the present paper that studied the conceptual opportunities that are made possible by the ODE-related formulation of constrained optimization problems.

### 4.3 Example 3: Solving Nonlinear Mixed-Integer Optimization Problems

In this section, presentation is done for the special case where all the decision variables are binary. The case where some decision variable can be continuous can be obtained easily with extra notational complexity.

Consider the optimization problem given by:

$$\min_{x \in \mathbb{R}^n} f(x) \quad \text{under } \bar{c}_i(x) \leq 0 \text{ and } x_i \in \{0, 1\} \quad (50)$$

for all  $i \in \{1, \dots, \bar{n}_c\}$ .

It is well known that this problem can be put in the standard form (1) by transforming the binary constraints  $x_i \in \{0, 1\}$  into standard constraints of the form

$$x_i - x_i^2 \leq 0 \quad -x_i \leq 0 \quad \text{and} \quad x_i - 1 \leq 0 \quad (51)$$

which yields a number of inequality constraints  $n_c = \bar{n}_c + 3n$ . Moreover, the integer  $n_\psi$  that characterizes the growth of  $\psi$  [see (11) and (12)] is given by  $n_\psi = 2m$  where  $m$  is the exponent used in the definition (3) of  $\psi$ .

Using  $m = 2$  leads to the following definition of  $\psi(x)$ :

$$\begin{aligned} \psi(x) &= \sum_{i=1}^{n_c} [\max\{0, c_i(x)\}]^2 + \sum_{i=1}^n [\max\{0, x_i - x_i^2\}]^2 \\ &+ \sum_{i=1}^n [\max\{0, -x_i\}]^2 + \sum_{i=1}^n [\max\{0, x_i - 1\}]^2 \end{aligned}$$

Now applying the result of Proposition 1 suggests the solution of the combinatoric optimization problem can be done by integrating the following ODE:

$$\dot{x} = - \left[ \sum_{i=1}^4 \frac{(\lambda \cdot g(x, \rho))^{i-1}}{(i-1)!} \right] \times \bar{f}_x(x, \rho) \quad (52)$$

$$\dot{\rho} = \gamma \times \psi(x) \quad (53)$$

Now obviously the admissible set is not a convex set and the presence of local minima is very likely. The following algorithm can be used to visit such local minima successively. In this algorithm a successively modified cost function  $f^{(s)}(\cdot)$  where  $f^{(0)} \equiv f$  is initialized to the original cost and where  $s$  denotes the number of already visited local minima. The local minimum  $x^{(s)}$  is found by integrating the ODE defined by the weighted function  $\bar{f}^{(s)}(x, \rho)$  and its corresponding norm of the gradient  $g^{(s)}(x, \rho) := \bar{f}_x^{(s)}(x, \rho)$ , namely:

$$\dot{x} = - \left[ \sum_{i=1}^4 \frac{(\lambda \cdot g^{(s)}(x, \rho))^{i-1}}{(i-1)!} \right] \times \bar{f}_x^{(s)}(x, \rho) \quad (54)$$

$$\dot{\rho} = \gamma \times \psi(x) \quad (55)$$

This is done starting from the initial condition  $(x^{(s-1)}, 0)$  and integrating the ODE until some stopping conditions on both  $\psi$  and  $g$  are satisfied. Then a term is added to the cost function which makes the current solution  $x^{(s)}$  inappropriate. This can be done by first defining a neighbor vector  $z^{(s)}$  to  $x^{(s)}$  such that:

$$\|z^{(s)} - x^{(s)}\|_\infty = 1 \quad \text{and} \quad c(z^{(s)}) \leq 0 \quad (56)$$

The new cost function  $f^{(s+1)}$  is now defined by:

$$f^{(s+1)}(x) := f^{(s)}(x) + (1 + 2f^{(s)}(z^{(s)})) \cdot \exp\left(\frac{\mu}{4} \|x - x^{(s)}\|^2\right)$$

Now for sufficiently high  $\mu$ , this new cost function is such that  $x^{(s)}$  is no more a local minimum since

$$\bar{f}^{(s+1)}(x^{(s)}) = f^{(s)}(x^{(s)}) + f^{(s)}(z^{(s)}) + 1 > f^{(s)}(z^{(s)})$$

therefore, incrementing  $s$  and firing the integration of the new resulting ODE (54)-(55) starting from the initial condition  $(x^{(s)}, 0)$  leads to a necessarily different minimum and so on.

The only assumption that is implicitly assumed is that there always exists a neighbor vector  $z^{(i)}$  that is admissible in the sense of (56). If this is not satisfied, less close  $z^{(s)}$  can be searched provided that a deterministic generation process is defined.

## 5 Conclusion and Future Work

In this paper, it is shown that the solution of optimization problems with inequality constraints can be obtained by solving appropriately defined ODEs. In these ODEs, simultaneous dynamics are given to the decision variable as well as to the weight associated to the exact penalty term on the constraint violation. One of the major impacts of this result lies in the possibility to design analogic circuits that can quickly and physically integrate the corresponding ODEs. Pushing this latter idea towards a concrete realization is the obvious follow up of the present work. Another direction is to use the result to derive a fast gradient algorithm together with its associated certification bounds regarding the number of iterations that would be necessary to achieve a prescribed level of precision following the steps of [12] while including affine constraints that are not considered in [12]. This was not possible precisely because when standard fast gradient is used, only projection on the box-like set can be done while guaranteeing the decrease of the cost function. The formulation proposed in the present paper provide generalization of this property to an extended monotonically decreasing cost function provided that the r.h.s of the ODE is used as an extended gradient.

## References

- [1] E. Birgin and J. M. Martinez. Augmented lagrangian method with non monotone penalty parameters for constrained optimization. *Computational Optimization and Applications*, 51(3):941–965, 2012.
- [2] R. Byrd, M. Hribar, and J. Nocedal. An interior point algorithm for large-scale nonlinear programming. *SIAM Journal on Optimization*, 9(4):877–900, 1999.
- [3] Level-Shifting Multiple-Input Switched-Capacitor Voltage Copier. Y. yuanmao and k. w. e. cheng. *IEEE Transaction on Power Electronics*, 27(2):828–837, 2012.
- [4] I. M. Filanovski and H. P. Baltes. Simple cmos analog square-rooting and squaring circuits. *IEEE Transactions on Circuits and Systems - I Fundamental Theory and Applications*, 39(4):312–315, 1992.
- [5] E. O. Gunes and F. Anday. Realization of voltage and current-mode transfer functions using unity gain cells. *International Journal of Electronics*, 83(2):209–213, 1997.
- [6] R. Hidayat, K. Dejhan, P. Moungnoul, and Y. Miyanaga. Ota-based high frequency CMOS multiplier and squaring circuit. In *Proceedings of the 2008 International Symposium on Intelligent Signal Processing and Communication Systems*, 2008.
- [7] Jorge R. Leis and Mark A. Kramer. Algorithm 658: Odessa—an ordinary differential equation solver with explicit simultaneous sensitivity analysis. *ACM Trans. Math. Softw.*, 14(1):61–67, March 1988.
- [8] D. Q. Mayne, J. B. Rawlings, C. V. Rao, and P. O. Scokaert. Constrained model predictive control: Stability and optimality. 36:789–814, 2000.
- [9] Y. Nesterov. A method of solving a convex programming problem with convergence rate  $o(1/k^2)$ . *Soviet Mathematics Doklady*, 27(2):372–376, 1983.
- [10] Y. Nesterov. *Introductory Lectures on Convex Optimization*. Springer-verlag, Berlin, Germany, 2004.
- [11] C. A. Papazoglou and C. A. Karybakas. Noninteracting electronically tunable CCII-based current-mode biquadratic filters. *IEE Proc. Circuits Devices Syst.*, 144(3):178–184, 1997.
- [12] S. Richter, C.N. Jones, and M. Morari. Computational complexity certification for real-time mpc with input constraints based on the fast gradient method. *Automatic Control, IEEE Transactions on*, 57(6):1391–1403, June 2012.
- [13] D. Voss and S. Abbas. Block predictor-corrector schemes for the parallel solution of ODEs. *Computers & Mathematics with Applications*, 33(6):65 – 72, 1997.